



TITLE:

Brachyury, Foxa2 and the cis-Regulatory Origins of the Notochord

AUTHOR(S):

José-Edwards, Diana S.; Oda-Ishii, Izumi; Kugler, Jamie E.; Passamaneck, Yale J.; Katikala, Lavanya; Nibu, Yutaka; Di Gregorio, Anna

CITATION:

José-Edwards, Diana S. ...[et al]. Brachyury, Foxa2 and the cis-Regulatory Origins of the Notochord. PLOS Genetics 2015, 11(12): e1005730.

ISSUE DATE:

2015-12-18

URL:

<http://hdl.handle.net/2433/214348>

RIGHT:

© 2015 José-Edwards et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

RESEARCH ARTICLE

Brachyury, Foxa2 and the *cis*-Regulatory Origins of the Notochord

Diana S. José-Edwards, Izumi Oda-Ishii^a, Jamie E. Kugler, Yale J. Passamaneck^{ab}, Lavanya Katikala, Yutaka Nibu, Anna Di Gregorio^{ac*}

Department of Cell and Developmental Biology, Weill Medical College of Cornell University, New York, New York, United States of America

^a Current address: Division of Biological Science, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake, Sakyo, Kyoto, Japan

^b Current address: Kewalo Marine Laboratory, University of Hawaii, Honolulu, Hawaii United States of America

^c Current address: Department of Basic Science and Craniofacial Biology, College of Dentistry, New York University, New York, New York, United States of America

* adg13@nyu.edu



OPEN ACCESS

Citation: José-Edwards DS, Oda-Ishii I, Kugler JE, Passamaneck YJ, Katikala L, Nibu Y, et al. (2015) Brachyury, Foxa2 and the *cis*-Regulatory Origins of the Notochord. PLoS Genet 11(12): e1005730. doi:10.1371/journal.pgen.1005730

Editor: Susan E. Mango, Harvard University, UNITED STATES

Received: August 19, 2015

Accepted: November 16, 2015

Published: December 18, 2015

Copyright: © 2015 José-Edwards et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Grant GM100466 from the National Institute of General Medical Sciences, <http://www.nigms.nih.gov/Pages/default.aspx>, Grant 1-FY11-468 from the March of Dimes Foundation, <http://www.marchofdimes.org>, Uehara Memorial Foundation of Japan <http://www.ueharazaidan.or.jp> and by the Alice Bohmfalk Charitable Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

A main challenge of modern biology is to understand how specific constellations of genes are activated to differentiate cells and give rise to distinct tissues. This study focuses on elucidating how gene expression is initiated in the notochord, an axial structure that provides support and patterning signals to embryos of humans and all other chordates. Although numerous notochord genes have been identified, the regulatory DNAs that orchestrate development and propel evolution of this structure by eliciting notochord gene expression remain mostly uncharted, and the information on their configuration and recurrence is still quite fragmentary. Here we used the simple chordate *Ciona* for a systematic analysis of notochord *cis*-regulatory modules (CRMs), and investigated their composition, architectural constraints, predictive ability and evolutionary conservation. We found that most *Ciona* notochord CRMs relied upon variable combinations of binding sites for the transcription factors Brachyury and/or Foxa2, which can act either synergistically or independently from one another. Notably, one of these CRMs contains a Brachyury binding site juxtaposed to an (AC) microsatellite, an unusual arrangement also found in Brachyury-bound regulatory regions in mouse. In contrast, different subsets of CRMs relied upon binding sites for transcription factors of widely diverse families. Surprisingly, we found that neither intra-genomic nor interspecific conservation of binding sites were reliably predictive hallmarks of notochord CRMs. We propose that rather than obeying a rigid sequence-based *cis*-regulatory code, most notochord CRMs are rather unique. Yet, this study uncovered essential elements recurrently used by divergent chordates as basic building blocks for notochord CRMs.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Transcription factors control the spatial and temporal expression of a multitude of genes by binding their *cis*-regulatory modules (CRMs). In this study, we investigated the architecture and composition of CRMs that direct gene expression in the notochord, a structure necessary for the support and patterning of the embryonic body plan of all chordates. We used the simple chordate *Ciona* to carry out a comparative study of notochord CRMs and we identified the sequences necessary for their function. These sequences, in turn, highlighted the existence of multiple mechanisms that enable gene expression in the notochord. Surprisingly, combinations of binding sites identical to those found in active CRMs were not necessarily able to direct notochord gene expression and were often poorly conserved among congener species. These results challenge the concept of a notochord-specific *cis*-regulatory “code”, and outline the limitations of methods for CRM identification that rely upon interspecific conservation of non-coding sequences. Nevertheless, a broad comparison of the structure of the *Ciona* CRMs with that of the notochord CRMs characterized thus far from all chordates outlines the existence of essential evolutionarily conserved building blocks, such as binding sites for the transcription factors Brachyury and Foxa2, that are shared by subsets of these regulatory modules.

Introduction

Cis-regulatory modules (CRMs), or enhancers, are genomic DNA regions that dictate location, timing and rate at which one or more genes are expressed [1]. These regions have variable length and contain a flexible number of binding sites for transcription factors that function as either activators or repressors [2]. Point mutations in one or more of the functional binding sites within a CRM can alter its spatial and temporal properties, or cause its partial or complete inactivation. Recent estimates suggest that the human genome contains hundreds of thousands of CRMs that are believed to be mainly responsible for the developmental and functional complexity of different cells, tissues, and organs [3]. Notably, mutations and deletions of human enhancers have been associated with developmental defects, disease, and cancer [4–6]. However, in the human genome, as well as in several others, CRMs can be located up to thousands of kilobases away from the genes that they control and are brought closer to their target promoters after being bound by specialized proteins that bend the DNA [7]. Furthermore, CRMs can be located within introns and/or other untranslated regions [8], or can be grouped into synergistically acting clusters called super-enhancers [9]. The crucial roles of CRMs, their complexity and their elusive nature, render a *cis*-regulatory code a highly desirable tool that would greatly simplify the genome-wide identification of CRMs with related properties. Studies aimed at identifying tissue-specific *cis*-regulatory codes have focused on genome-wide searches of clusters of known transcription factor binding sites [10] and on interspecific conservation of clusters of binding sites and/or larger non-coding sequences [11]. Nevertheless, recent research suggests that conserved clusters of binding sites are often non-functional [12] and that even evolutionarily ultraconserved genomic regions do not necessarily possess *cis*-regulatory activity [13].

The aim of the present study was to determine the structure and the functional binding sites of CRMs that shared comparable *cis*-regulatory activity and were presumably co-regulated, and to look for elements that could define a tissue-specific *cis*-regulatory code. We centered our analysis on CRMs active in the notochord, the most distinctive of chordate synapomorphies [14,15]. In all chordates, the notochord is the main source of support for the developing

embryo and an essential patterning center for many of its structures and organs [16]. In vertebrates, the notochord is replaced by the vertebral column and its remnants form the *nuclei pulposi* of the intervertebral discs [17]. For the present study we used as a model system the tunicate *Ciona*, an invertebrate chordate that couples a compact, fully annotated genome with ease of transgenesis and tractable notochord [18,19]. According to phylogenomics data, tunicates are the invertebrate chordates most closely related to vertebrates [20], and thus provide an opportunity to reconstruct the genetic circuitry and the evolutionary origins of the notochord through the identification of *cis*-regulatory sequences that enable gene expression in this structure [21–23].

We began this analysis with the characterization of fourteen notochord CRMs from *Ciona*. After isolating the minimal sequences necessary for their function, we tested whether these minimal sequences could be used to predict related notochord CRMs. We also evaluated the evolutionary conservation of CRM sequences between two *Ciona* species, *C. intestinalis* and *C. savignyi*, and compared the structure of the *Ciona* notochord CRMs to fully characterized notochord CRMs from other chordates, including mouse and zebrafish.

Rather than a *sensu stricto cis*-regulatory code, this study elucidated various combinations of functional transcription factor binding sites that function in a context-dependent fashion. These binding sites are often poorly conserved interspecifically, and therefore would have been missed by conservation-based methods of enhancer detection. However, despite the intraspecific and interspecific variability in their composition and function, binding sites for Brachyury and Foxa2 emerged as recurrent hallmarks of notochord CRMs from highly divergent chordates.

Results and Discussion

We identified fourteen CRMs that can induce gene expression in the *Ciona* notochord. To avoid sequence and/or positional biases, all but one of the notochord CRMs (Fig 1) were isolated through testing of random genomic regions (S1 Table). Minimal notochord enhancers spanning 80–547 bp were subsequently identified through sequence-unbiased truncation analyses, involving *in vivo* testing of ~200 constructs (S1, S2 and S3 Figs). Lastly, we assessed the effects of site-directed mutations targeting either known putative transcription factor (TF) binding sites or uncharacterized sequences. The results of these studies are condensed in Fig 1.

We found that the majority of the CRMs (9/14, 64.3%) require binding sites for the TFs *Ciona* Brachyury (Ci-Bra) and/or Ci-FoxA-a (Foxa2/fkh/HNF3beta ortholog; hereinafter Ci-Fox); in contrast, binding sites for TFs of widely different families were responsible for the function of the remaining five notochord CRMs. This analysis also revealed unexpected characteristics of these regulatory elements. For instance, enrichment for a particular binding site was not a reliable predictor of either functionality or cooperativity (e.g., all Ci-Fox sites in Ci-CRM70 are dispensable; Figs 1 and S1). In some instances, only one of the multiple copies/variants of a given TF binding site was required for notochord gene expression (e.g., only one of the seven Ci-Bra sites in Ci-CRM99 is necessary; Figs 1 and S3). Furthermore, even CRMs necessitating the same types of binding sites could function differently: a Myb-like site worked individually in one CRM (*Ci-C6ST-like7*), and in combination with a related Myb-like site in another (Ci-CRM76) (Figs 1 and S1).

We had previously described a notochord CRM, associated with the gene *Ci-tune*, activated by synergistic Ci-Bra and Ci-Fox binding sites [24]. In this study, we found that Ci-CRM96 relies on the same type of synergism (Fig 2A), and although the sequences of the Ci-Bra and Ci-Fox sites differ between these two CRMs, their spacing is comparable (48 bp in Ci-CRM96, 46 bp in *Ci-tune*). In contrast, the multiple Ci-Bra and Ci-Fox sites in Ci-CRM24 act

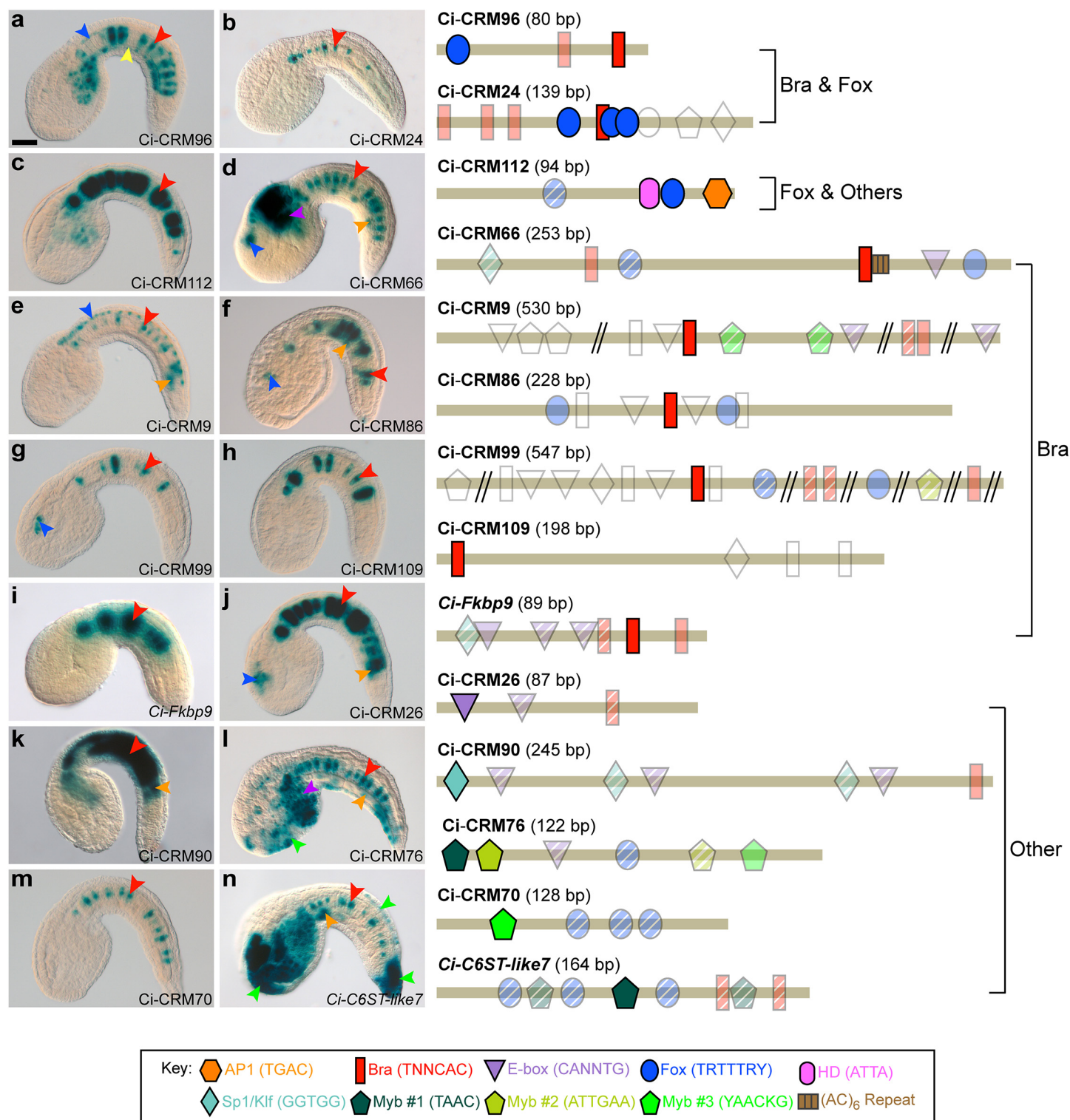


Fig 1. A comparative study of notochord CRMs in *Ciona*. **a-n:** Microphotographs of transgenic *Ciona* embryos expressing the *LacZ* reporter in the notochord (red arrowheads) under the control of 14 CRMs. (Right) Schematic representations of the 14 minimal notochord CRMs. Putative transcription factor binding sites are mapped along the length of each enhancer (tan bar), as indicated in the key (bottom). Point mutations uncovered site(s) required for notochord expression (colored and opaque) as well as sites that did not evidently contribute (colored, but transparent). Putative binding sites deemed dispensable through truncations are colored and hatched. Untested putative sites are outlined in gray. Additional staining domains are indicated by arrowheads, colored as follows: blue: CNS, yellow: endoderm, orange: muscle, purple: mesenchyme, green: epidermis. Embryos are oriented with dorsal up and anterior to the left. Scale bar: 40 μ m. See also [S1](#), [S2](#) and [S3](#) Figs

doi:10.1371/journal.pgen.1005730.g001

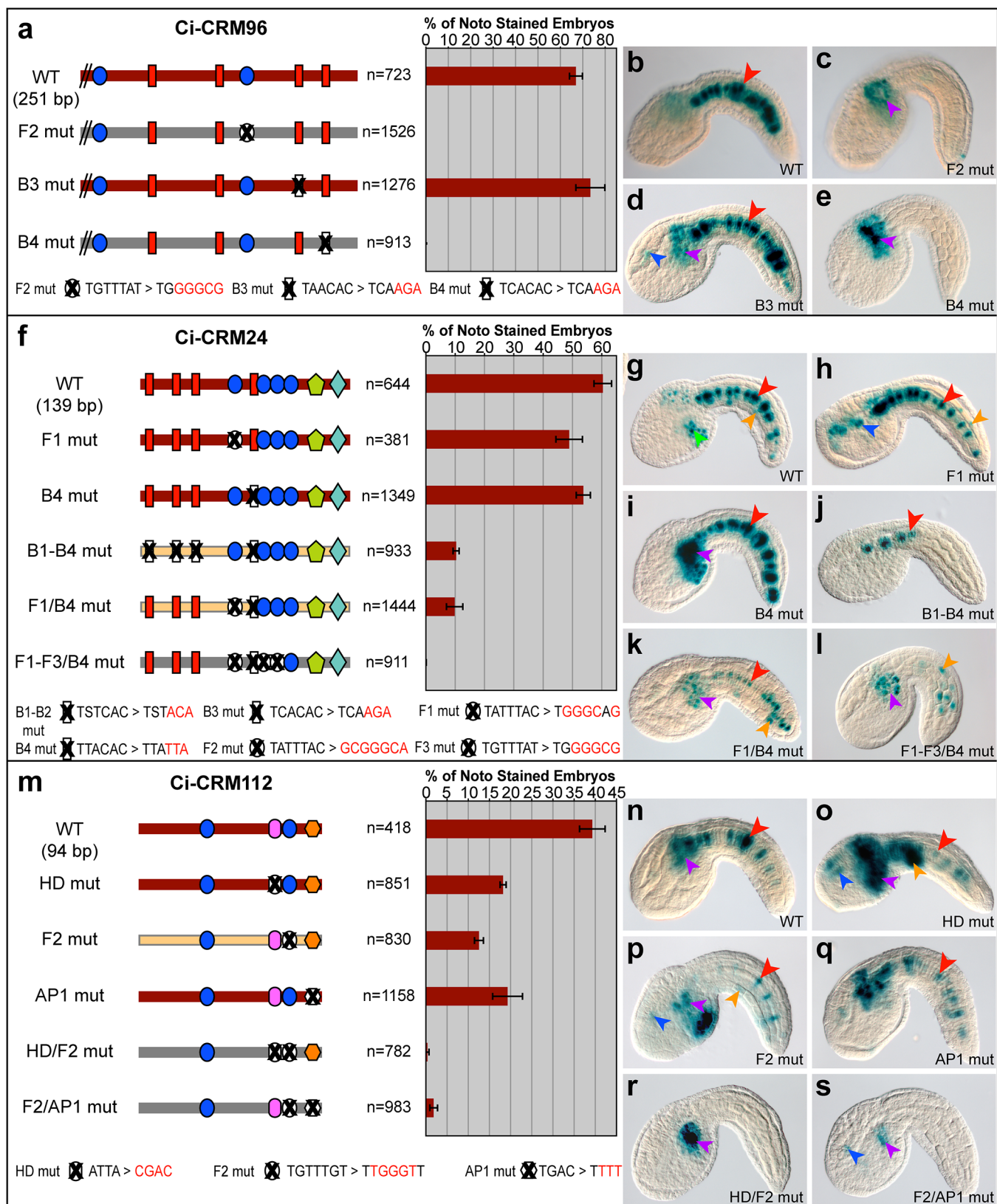


Fig 2. Alternative regulatory mechanisms of notochord CRMs requiring Ci-Bra and/or Ci-Fox binding sites. **a,f,m:** (Left) Schematic representations of wild-type (WT) and site-directed mutant CRMs; TF binding sites are as in Fig 1, with the mutant sequences indicated at the bottom of each panel. Mutated binding sites are colored in white and covered by “X” signs. Maroon bars represent constructs able to elicit notochord expression, while configurations exhibiting weak or no notochord staining are depicted by yellow and gray bars, respectively. (Right) Quantification of the fraction of the total stained embryos showing notochord expression after electroporation of the constructs at the left of each bar. n: number of fully developed stained embryos. Error bars denote standard deviation from the mean. **b-e, g-l, n-s:** Microphotographs of embryos expressing the transgenes indicated at the bottom right of each panel. Arrowheads are color-coded as in Fig 1. Abbreviations: WT: wild-type, F: Fox binding site, B: Brachyury binding site, HD: homeodomain, AP1: activator protein 1, mut: mutated, noto: notochord. In f, “S” stands for C/G. See also S2 Fig.

doi:10.1371/journal.pgen.1005730.g002

redundantly, as individual mutations (e.g., Fox1 and Bra4, Fig 2F) are not detrimental to notochord staining (Fig 2F–2I), and reduction/loss of notochord staining is only obtained through compound mutations (Fig 2F, 2J, 2K and 2L). Unlike the previous CRMs, Ci-CRM112 is devoid of Ci-Bra sites (Fig 2M). In this case, putative homeodomain (HD) and activator protein 1 (AP1) sites appear to work cooperatively with a Ci-Fox site, since all single mutations decrease notochord staining (Fig 2M–2Q), and simultaneous mutations of the functional Ci-Fox site and either the HD or AP1 sequences result in loss of staining (Figs 2M, 2R, 2S and S2).

Six CRMs rely on individual Ci-Bra binding sites (Figs 1, S1 and S3). Counterintuitively, the sequences of indispensable Ci-Bra sites differ for each Ci-Bra-dependent CRM, and sites with identical core sequences may be necessary in one context, but not in another (e.g., the TTGCAC sites in Ci-CRM109 and *Ci-Fkbp9*; S1 and S3 Figs). To uncover the molecular foundations of such differences, we assessed the roles of sequences directly adjacent to the necessary Ci-Bra binding sites. For Ci-CRM66, which lies within an intron of *Ci-Ephrin3*, we found that mutation of a single Ci-Bra binding site drastically decreased, but did not abolish, notochord staining (Figs 3A, 3E, 3J and S3). Linker-scanning mutagenesis revealed that the most detrimental mutations were those affecting an (AC)₆ microsatellite [25] directly abutting the TCACAC Ci-Bra site (Fig 3B). Mutation of the first two (AC) pairs (Fig 3C) caused a sharp drop in notochord expression (Fig 3H and 3J), as did a mutation that caused a “frame-shift” of the microsatellite sequence (Fig 3B and 3F), suggesting that uninterrupted periodicity between the Ci-Bra binding site and this sequence may be required for the function of this CRM. The number of intact repeats also influenced activity (Fig 3B), and the mutation of the entire microsatellite abolished notochord expression (Fig 3C, 3I and 3J). Notably, ChIP-chip studies of genomic targets of Brachyury in differentiating mouse embryonic stem cells showed that this TF often binds (AC) repeats [26]. The *Ciona intestinalis* genome contains only nine copies of an (AC)_{≥6} microsatellite abutting a TCACAC Ci-Bra binding site; however, despite their reported occupancy by Ci-Bra in early embryos [27], none of the remaining eight regions directed notochord gene expression (S2 Table).

We also searched the sequences of the remaining five CRMs that rely on single Ci-Bra binding sites for clues on the mechanisms that might create the appropriate context for their function. Even though mouse Brachyury was initially found to bind the palindromic sequence T(G/C)ACACCTAGGTGTGA [28], it was later shown that TNNCAC core half-sites are efficiently bound by Brachyury proteins from mouse and other organisms, including *Ciona* [29–32]. Our results confirm that a palindromic organization is not required; instead, we observed that 50% of the required Ci-Bra sites matched either the TNNCACCTAM or the CTAMGTGNNAC consensus (core sites underlined) (Fig 3K). Consequently, we selectively mutated the adjacent nucleotides while leaving the TNNCAC cores intact and found that in the case of Ci-CRM109 and Ci-CRM99 disruption of the CTAM sequence had the same effect as the mutation of the cores (Fig 3L–3S). Similar results were obtained through the mutation of this stretch in the *Ci-ABCC10* CRM [33]. In contrast, mutation of the CTAM sequence within Ci-CRM86 left notochord staining unaffected (Fig 3T–3W) and a CTAM-containing Ci-Bra binding site within Ci-CRM9 was found to be dispensable (S3 Fig). We conclude that the CTAM extension is not

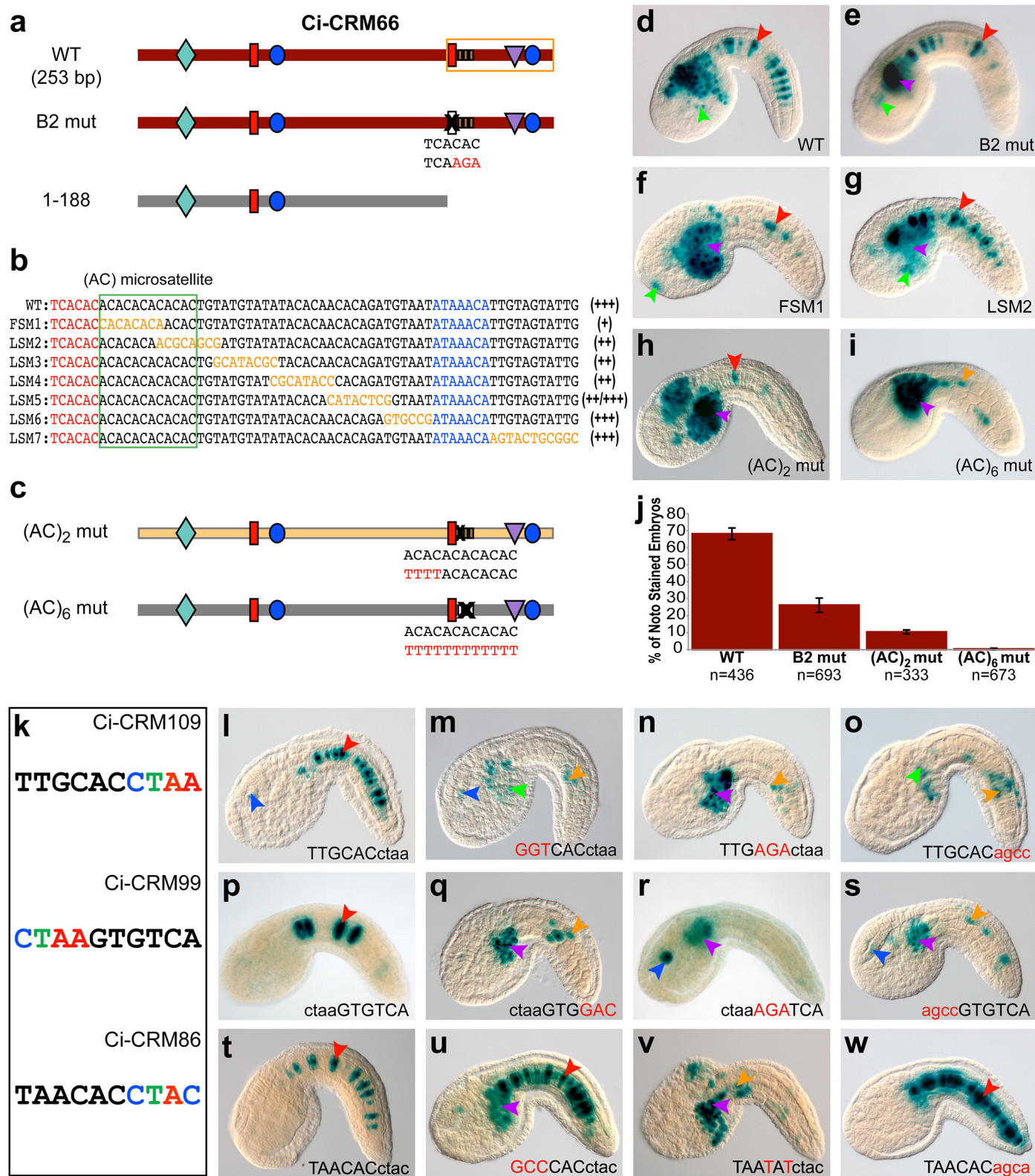


Fig 3. The function of individual Ci-Bra binding sites can be modulated by either an (AC) microsatellite or a flanking sequence. a,c: Schematic representations of wild-type (WT) and mutant CRMs, as described and colored in [Fig 2](#); the (AC) microsatellite sequence is schematized as a segmented brown rectangle. b: Mutational series of the area boxed in orange in the 253-bp construct. Red and blue nucleotides correspond to the Ci-Bra and Ci-Fox

sites, respectively, and orange nucleotides indicate the bases changed in each mutant plasmid. The (AC)₆ microsatellite sequence is boxed in green. The relative ability of each construct to direct notochord gene expression is shown by plus signs at the right of each sequence. **d-i**: Photos of embryos electroporated with the constructs depicted in **a,b,c**; arrowheads are color-coded as in **Fig 1**. **j**: Quantification of notochord-stained embryos harboring the constructs in **a,c**. Error bars indicate standard deviation from the mean. **k**: Identification of an extended CTAM sequence (colored) shared by a subset of individually-acting Ci-Bra binding sites. **l-w**: Microphotographs of embryos carrying wild-type CRMs (**l,p,t**) compared to embryos carrying various mutant versions of Ci-CRM109 (**m-o**) Ci-CRM99 (**q-s**) and Ci-CRM86 (**u-w**). Core Ci-Bra binding sites are capitalized. Mutations are depicted in red. Abbreviations: FSM: “frame-shift” mutation, LSM: linker scanning mutation. See also **S3 Fig**.

doi:10.1371/journal.pgen.1005730.g003

entirely predictive of whether a CRM will necessitate a single Ci-Bra site, and the binding sites that possess it are not always necessary. It is also conceivable that a fraction of the binding sites that we tentatively attributed to Ci-Bra might be interchangeably or exclusively utilized by Ci-Tbx2/3, the only other T-box protein present in the *Ciona* notochord, which acts as a mediator of Ci-Bra [34]. The sequences flanking the core TNNCAC site might therefore be required for binding specificity of either T-box factor, Ci-Bra or Ci-Tbx2/3.

In the last group of five minimal CRMs, the sequences required for notochord expression were neither Ci-Bra nor Ci-Fox binding sites (**Fig 1**), but instead resembled sites for bHLH (Ci-CRM26), Klf/Sp1 (Ci-CRM90), and Myb-like factors (Ci-CRM70, Ci-CRM76 and *Ci-C6ST-like7*) (**S1 Fig**). These results are consistent with previous reports of notochord-expressed bHLH, Klf6 and Klf15 TFs [35–37], and of a *Myb*-related gene in *Ciona* [38]. The requirement for two short Myb-like sites in Ci-CRM76 (**Fig 1**) led us to hypothesize that its activity might require a specific architecture. Accordingly, we found that while reversing the orientation of one of the Myb-like sites (abbreviated as “M”), M2-2, had no effect, transposing the order of the two required Myb-like sites, M1-5 and M2-2, largely decreased notochord staining (**S4 Fig**). Furthermore, increasing the spacing between M1-5 and M2-2 (4 bp) to that of the dispensable sites, M2-1 and M1-4 (8 bp), caused an even more substantial reduction of reporter gene expression in the notochord (**S4 Fig**). Nevertheless, seven genomic regions containing Myb-like sites with the identical composition, orientation and spacing as Ci-CRM76, all of which mapped near notochord genes, did not yield detectable notochord expression when tested *in vivo* (**S3 Table**).

Additional sequence inspection identified non-microsatellite repeats in various CRMs. Combinations of recurring motifs and/or evolutionarily conserved TF binding sites have guided the identification of CRMs active in the *Ciona* muscle [21,39–42] and central nervous system (CNS) [41,43], as well as in various tissues/embryonic territories of *Drosophila* [10,44,45] and in the zebrafish notochord [46]. For these reasons, we sought to investigate whether these repeats could aid in the prediction of novel notochord CRMs in *Ciona intestinalis*. We noticed that Ci-CRM90 features two nearly identical 73-bp sequence blocks, each containing two copies of a smaller 20-bp repeat; moreover, a sequence motif related to the 20-bp repeat was found in Ci-CRM9 (**S4 Fig**). Ci-CRM26 contains a 19-bp tandem repeat, whose first copy overlaps with the E-box required for activity. The exact sequences of both of these repeats are unique in the *Ciona intestinalis* genome; however, shorter variations of the Ci-CRM26 repeat are seen in four other notochord CRMs (**S4 Fig**). To assess the predictive ability of functional binding sites and motifs, we tested 36 genomic fragments containing arrangements of binding sites and/or motifs identical or similar to those found in the Ci-CRMs (**Fig 1**). We only detected notochord expression in one construct (**S3 Table**, **S4 Table**): the short motif found in Ci-CRM26, which occurs ~3,017 times in the *Ciona intestinalis* genome, led us to the identification of a novel notochord CRM within the *Ci-Noto2* locus (**S4 Fig** and **S4 Table**).

We also tested whether interspecific sequence homology could improve the prediction of notochord CRMs, since evolutionary conservation is widely used to pinpoint *Ciona cis*-regulatory regions (e.g., [47–49]). The CRMs presented here were isolated using a conservation-

independent approach, but when we retrospectively assessed this parameter, we observed surprising interspecific variability among their sequences. Indeed, many of these *Ciona intestinalis* CRMs display limited conservation, if any, with *Ciona savignyi* (S4 Fig). In addition, even though some binding sites, such as the Ci-Fox and E-box sites of Ci-CRM76, are perfectly conserved between the two *Ciona* species, neither is required for activity (S4 Fig); this suggests that even interspecifically conserved notochord TF binding sites are not reliable indicators of functionality. These results concur with studies in *Drosophila* that suggest that clustered binding sites within CRMs might be retained over evolution for reasons other than selection or functional necessity [12].

In sum, the unexpected variety and flexibility of the mechanisms that we have described here limited our ability to predict notochord CRMs from sequence alone. Yet, although our results seem to question the existence of a straightforward notochord *cis*-regulatory code, this study uncovered recurring grammatical elements shared by notochord CRMs. In particular, Brachyury and Foxa2 binding sites emerge as the basic building blocks of most *Ciona* notochord CRMs (Fig 4A), and these results are consistent with findings in other chordates. In fact, Brachyury binding sites have been found to be critical for the function of notochord in different animals (e.g. [29,50]), and our previous studies in *Ciona* show that they can act either individually or cooperatively [33,34,53]. Their association with (AC) microsatellites in Ci-CRM66 and in the mouse genome [26] might represent a recurring feature of a distinct class of notochord CRMs (Fig 4A). Foxa2 sites are required in notochord CRMs from zebrafish and mice [46,54], although they are rarely sufficient to initiate expression when in single copy, and often necessitate additional sequences [46,58,61] whose identity appears to be lineage-specific (Fig 4A and 4C). These observations and our previous results [33] reflect the reported pioneer chromatin-opening ability of Fox proteins [62], which may not be able to activate gene expression *per se* but are required to increase the accessibility of CRMs to other transcription factors, such as Brachyury and/or other notochord-specific activators.

The basic *cis*-regulatory repertoire that we have uncovered was likely expanded *via* vertebrate-specific evolutionary events; such events include the notochord deployment of additional TFs, such as homeobox and Hox proteins and their co-factors, which are remarkably underrepresented in the tunicate notochord, [63] along with the duplication and consequent divergence of regulatory regions.

Materials and Methods

Embryo culture, fixation, electroporation and staining

Adult *Ciona intestinalis* were purchased from Marine Research and Educational Products (M-REP; Carlsbad, CA) and kept in an aquarium in recirculating artificial sea water at 17–18°C. Culturing and electroporations were carried out as previously described [64]. After electroporation, transgenic embryos were fixed in 0.2% glutaraldehyde and stained at 37°C with 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal) [64]. Stained embryos were washed in 500 μ L PBST (1X PBS, 0.1% Tween 20), post-fixed in 300–500 μ L of 4% paraformaldehyde in PBST, and stored at 4°C. To determine the comparative activities of wild-type and mutated constructs, the proportions of X-gal stained embryos exhibiting notochord staining were determined from at least three independent experiments. Data presented in graphs represent average values, with error bars denoting the standard deviation.

Plasmid construction

Genomic fragments for enhancer discovery and analyses were cloned into the pFBASP6 plasmid, which contains the *LacZ* reporter gene [64]. After the initial characterization of each

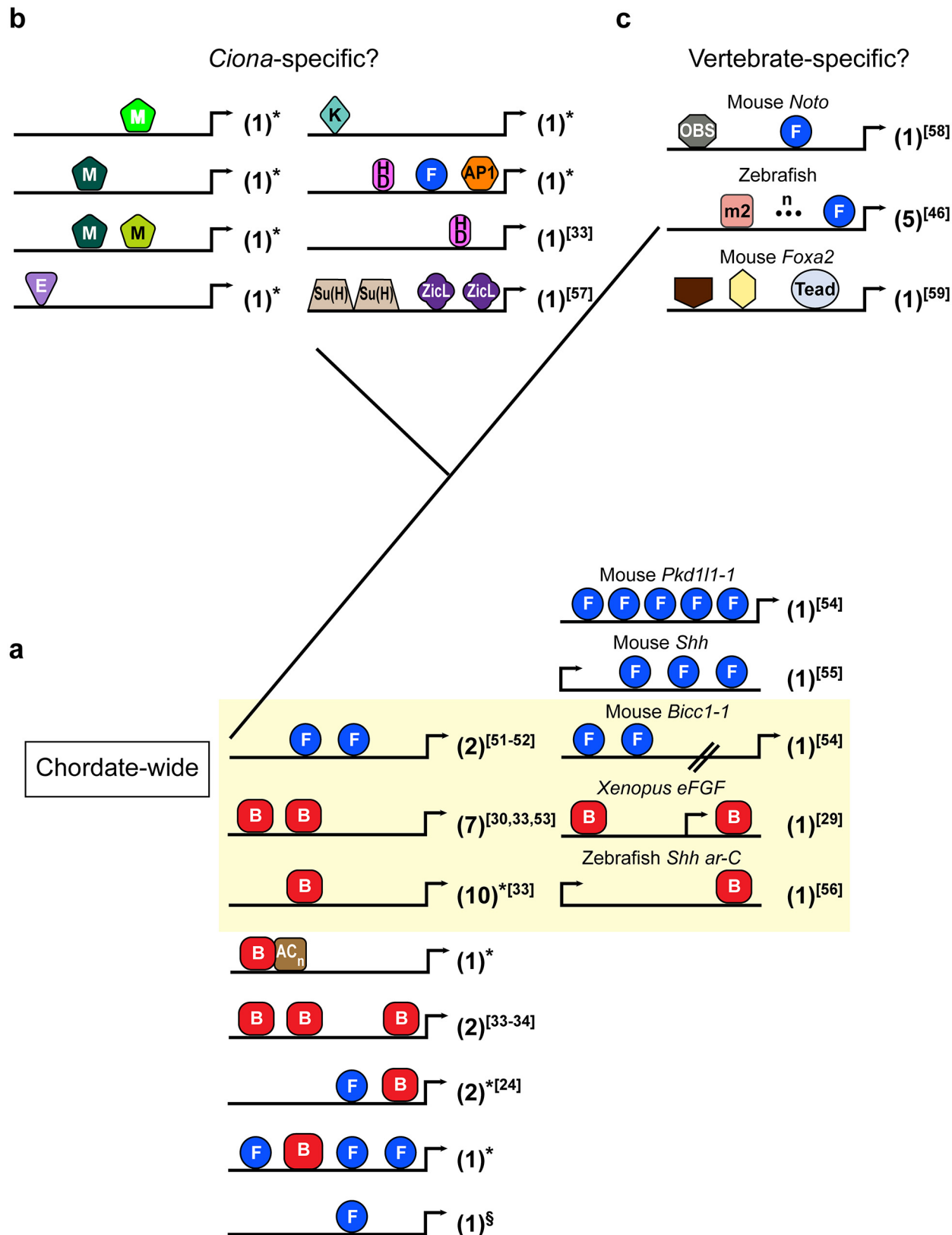


Fig 4. The *cis*-regulatory building blocks of notochord CRMs in chordate phylogeny. Schematic representation of all 46 experimentally validated and fully characterized notochord CRMs published in any chordate [24,29,30,33,34,46,51–59], grouped into 24 structural types. Among the 35 *Ciona* CRMs, 14 were described for the first time in this study (Fig 1). Notochord CRMs are symbolized by black lines, with arrows representing transcription start sites. Colored shapes depict putative transcription factor binding sites. Only experimentally validated binding sites required for the *in vivo* activity of each CRM are reported. The numbers in parentheses denote the number of related CRMs identified thus far that display each *cis*-regulatory arrangement. **a:** Chordate-wide *cis*-regulatory features of *Ciona* notochord CRMs (left column) and vertebrates (right column). The area highlighted in yellow encompasses notochord CRMs from *Ciona* and from vertebrates that show directly comparable binding sites and arrangements. Notochord CRMs above and below the yellow area rely on either reiterative or alternate configurations of Brachyury (B) and Foxa2 (F) functional binding sites. **b:** Notochord CRMs that, thus far, do not seem to have counterparts in other chordates, and are therefore tentatively classified as *Ciona*-specific. **c:** Notochord CRMs that currently do not appear to have counterparts in *Ciona* or other invertebrate chordates and are provisionally classified as vertebrate-specific. TF binding sites are abbreviated as follows: AP1: activator protein 1, B: Brachyury, F: Fox, E: E-box, HD: homeodomain, K: Krüppel-like, M, Myb-like, m2, Motif 2, OBS: orphan binding site. A brown pentagon and a yellow hexagon in the mouse *Foxa2* notochord CRM indicate required orphan binding sites. * this study; § notochord CRM associated with the *Ci-quaking* gene (KH.S115.4) [60].

doi:10.1371/journal.pgen.1005730.g004

notochord CRM, subsequent deletions and mutations were made either by utilizing unique restriction enzyme sites or by Polymerase Chain Reaction (PCR), using the smallest active DNA fragment as a template. A list of the oligonucleotides employed for PCR amplifications and the restriction sites used for cloning the most relevant constructs is provided in S5 Table.

For the predictions of notochord CRMs, suitable genomic regions were first identified by searching either the *Ciona* genome or a database of validated *Ciona* notochord genes for transcription factor binding sites, motifs or other sequence signatures present in notochord CRMs, using the GUFEE program [24]. Our database of *Ciona* notochord genes contained the sequences of the putative genomic loci of 300 notochord genes. We manually annotated the gene models from expression data present in the ANISEED database [38] and from our results. The sequences included in the database were extracted from the UCSC genome browser (*Ciona intestinalis* version 1) by Dr. John R. Edwards (Washington University, St. Louis).

Supporting Information

S1 Fig. Initial characterization of a subset of the notochord CRMs described in Fig 1. a–f: Schematic representations of wild-type notochord CRMs and site-specific mutants of selected binding sites (see Fig 1 for key). Maroon bars represent constructs capable of directing notochord expression of the *LacZ* reporter, while inactive configurations are depicted by gray bars. Mutagenized sites are colored in white and marked by “X” signs, and the mutant sequences are shown in red. Each panel contains microphotographs of representative transgenic embryos carrying selected plasmids. Colored arrowheads indicate stained domains as follows: red: notochord, blue: CNS, yellow: endoderm, orange: muscle, purple: mesenchyme, green: epidermis. Graphs display the percentage of embryos showing notochord staining among all stained embryos; error bars indicate the standard deviation. Abbreviations: B, Brachyury; E, E-box (presumptive bHLH binding site); F, Fox; HD, homeodomain; M, Myb-like; S/K1, Sp1/Klf (TIF)

S2 Fig. Deletion/mutation analysis of the notochord CRMs described in Fig 2. a,b,d: Schematic representations of wild-type notochord CRMs and site-directed mutants of the binding sites shown in the key (top right). The color-coding of the bars representing the DNA regions is the same as in Fig 2. “X” signs indicate mutagenized sites, and mutant sequences are shown in red. **c,e,f:** Representative embryos carrying a selection of the plasmids depicted in a,b,d. Colored arrowheads indicate stained domains as in S1 Fig. Abbreviations: AP1: Activator protein 1, Bra: Brachyury, HD: homeodomain. (TIF)

S3 Fig. Deletion/mutation analysis of the notochord CRMs described in Fig 3 and of Ci-CRM9. a–e: (Left) Schematic representations of notochord CRMs and their mutant versions.

Putative binding sites are depicted as shown in the key on the top right in **a**. Color-coding is as in **S1 Fig**. Mutagenized sites are indicated by “X” signs and the mutant sequences are in red. (Middle) Transgenic embryos carrying a selection of informative plasmids. Arrowheads are color-coded as in **S1 Fig**. Note that in **c** the B4 mutation was inserted in the 749-bp fragment since the minimal CRM (547-bp) exhibits a less consistent staining pattern. (Right, **b-e**) Quantification of notochord stained embryos harboring either wild-type or mutant constructs; error bars denote the SD. Abbreviations are as in **S1 Fig**. (TIF)

S4 Fig. Architectural constraints, sequence motifs and interspecies conservation of *Ciona* notochord CRMs. Related to Figs **1** and **3**. **a-d**: Impact of the alteration of structural features on the function of Ci-CRM76 in notochord cells. (Left) schematic representations of wild-type (WT) and mutant versions of Ci-CRM76 containing the changes in enhancer architecture highlighted in red. Putative Myb-like binding sites are named as in **S1 Fig**. Symbols for all other binding sites are as in **Fig 1**. The necessary Myb-like sites are marked by “1” and “2”. Arrows show the orientation of the binding sites of interest. (Right) Representative transgenic embryos obtained from the same batch of animals, harboring the plasmids summarized at their left. Arrowheads mark stained territories, as in **S1 Fig**. The percentage of embryos exhibiting notochord staining is reported in the lower right corner. M: Myb-like binding site. **e-g**: Sequence motifs shared by subsets of notochord CRMs. **e,f**: Schematic representations of notochord CRMs sharing distinctive sequence blocks. Tan bars symbolize notochord CRMs and diagonal parallel lines depict genomic regions that are present in the constructs but omitted from the figure for clarity. In Ci-CRM90, a 73-bp sequence, boxed in yellow, is imperfectly repeated in the 245-bp region shown here. Within this 73-bp sequence, four motifs were identified (#1–4) using the MEME software (<http://meme.nbcr.net>). A related motif was identified in the Ci-CRM9 sequence (boxed in yellow), adjacent to the Ci-Bra binding site necessary for its function. The sequences of all these motifs, and the derived consensus, are reported on the right. **f**: Another motif (light blue boxes) was found to be present in one or two copies in a different subset of CRMs. The sequences of its iterations, and the derived consensus, are reported on the right. The distances between the necessary site(s) and each motif are shown, unless they overlap. A closely related motif was found in Ci-CRM99. The CRMs included in this figure are depicted in a slightly different scale compared to the previous figures, to provide a more accurate representation of the distances among binding sites. **g**: Microphotograph of a transgenic embryo electroporated with the *Ci-Noto2* notochord CRM, which was predicted using the Ci-CRM26 motif. **h,i**: Variability in the interspecific conservation of notochord CRMs sequences between *Ciona intestinalis* and *Ciona savignyi*. (Top) VISTA plots (<http://pipeline.lbl.gov/cgi-bin/gateway2>) illustrating the sequence conservation across the “full-length” *Ci-Fkbp9* (**h**) and Ci-CRM76 (**i**) notochord CRMs between *Ciona intestinalis* (Ci) and *Ciona savignyi* (Cs), obtained utilizing the following parameters: calculation window, 80 bp; minimum conservation width, 50 bp; conservation identity, 70%. Conserved non-coding regions are depicted as pink peaks, conserved coding regions as blue peaks. The areas corresponding to the minimal CRMs identified and described in **Fig 1** are boxed in red. (Bottom) Sequence alignment of the Ci minimal notochord CRMs with the corresponding regions of Cs. In Ci, binding sites are highlighted as in **Fig 1**, whereas related non-syntenic putative binding sites, whenever present, are indicated in lighter colors in the Cs sequence.

(TIF)

S1 Table. Genomic locations of minimal notochord CRMs.

(DOCX)

S2 Table. Properties of (AC) microsatellite clusters tested for notochord activity.
(DOCX)

S3 Table. Properties of genomic regions near notochord genes showing arrangements of sites resembling those found in selected notochord CRMs.
(DOCX)

S4 Table. Properties of genomic regions containing sequence motifs found in subsets of notochord CRMs.
(DOCX)

S5 Table. Primers utilized for the PCR amplification of the most relevant constructs used for CRM characterization.
(DOCX)

Acknowledgments

We are thankful to Ms. Mami Takeda, Ms. Shruti Sharma, Ms. Karina Braslavskaya and Dr. Sara Peyrot for their precious technical help. We thank Dr. John R. Edwards (Washington University) for help and advice with the genomic searches and for critically reading the manuscript. We remain indebted to the late Dr. Eric Davidson for insightful discussion on *cis*-regulatory modules.

Author Contributions

Conceived and designed the experiments: ADG. Performed the experiments: DSJE IOI JEK YJP LK ADG. Analyzed the data: DSJE IOI JEK YJP LK YN ADG. Contributed reagents/materials/analysis tools: YN. Wrote the paper: DSJE ADG.

References

- Howard ML, Davidson EH. *Cis*-regulatory control circuits in development. *Dev Biol.* 2004; 271: 109–118. PMID: [15196954](#)
- Levine M. Transcriptional Enhancers in Animal Development and Evolution. *Curr Biol.* 2010; 20: R754–R763. doi: [10.1016/j.cub.2010.06.070](#) PMID: [20833320](#)
- Erokhin M, Vassetzky Y, Georgiev P, Chetverina D. Eukaryotic enhancers: common features, regulation, and participation in diseases. *Cell Mol Life Sci.* 2015; 72: 2361–2375. doi: [10.1007/s00018-015-1871-9](#) PMID: [25715743](#)
- Kleinjan D-J, Coutinho P. *Cis*-ruption mechanisms: disruption of *cis*-regulatory control as a cause of human genetic disease. *Brief Funct Genomic Proteomic.* 2009; 8: 317–332. doi: [10.1093/bfgp/elp022](#) PMID: [19596743](#)
- Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, et al. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science.* 2014; 346: 1373–1377. doi: [10.1126/science.1259037](#) PMID: [25394790](#)
- Melton C, Reuter JA, Spacek DV, Snyder M. Recurrent somatic mutations in regulatory regions of human cancer genomes. *Nat Genet.* 2015; 47: 710–716. doi: [10.1038/ng.3332](#) PMID: [26053494](#)
- Levine M, Cattoglio C, Tjian R. Looping Back to Leap Forward: Transcription Enters a New Era. *Cell.* 2014; 157: 13–25. doi: [10.1016/j.cell.2014.02.009](#) PMID: [24679523](#)
- Ott CJ, Blackledge NP, Kerschner JL, Leir S-H, Crawford GE, Cotton CU, et al. Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. *Proc Natl Acad Sci USA.* 2009; 106: 19934–19939. doi: [10.1073/pnas.0900946106](#) PMID: [19897727](#)
- Hnisz D, Schuijers J, Lin CY, Weintraub AS, Abraham BJ, Lee TI, et al. Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Mol Cell.* 2015; 58: 362–370. doi: [10.1016/j.molcel.2015.02.014](#) PMID: [25801169](#)

10. Markstein M, Markstein P, Markstein V, Levine MS. Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc Natl Acad Sci USA*. 2002; 99: 763–768. PMID: [11752406](#)
11. Poulin F, Nobrega MA, Plajzer-Frick I, Holt A, Afzal V, Rubin EM, et al. *In vivo* characterization of a vertebrate ultraconserved enhancer. *Genomics*. 2005; 85: 774–781. PMID: [15885503](#)
12. Lusk RW, Eisen MB. Evolutionary mirages: selection on binding site composition creates the illusion of conserved grammars in *Drosophila* enhancers. *PLoS Genetics*. 2010; 6: e1000829. doi: [10.1371/journal.pgen.1000829](#) PMID: [20107516](#)
13. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. 2006; 444: 499–502. PMID: [17086198](#)
14. Jiang D, Smith WC. Ascidian notochord morphogenesis. *Dev Dyn*. 2007; 236: 1748–1757. PMID: [17497687](#)
15. Satoh N, Tagawa K, Takahashi H. How was the notochord born? *Evol Dev*. 2012; 14: 56–75. doi: [10.1111/j.1525-142X.2011.00522.x](#) PMID: [23016975](#)
16. Stemple DL. Structure and function of the notochord: an essential organ for chordate development. *Development*. 2005; 132: 2503–2512. PMID: [15890825](#)
17. Lawson L, Harfe B. Notochord to Nucleus Pulposus Transition. *Curr Osteoporos Rep*. Springer US; 2015; 13: 336–341–341. doi: [10.1007/s11914-015-0284-x](#) PMID: [26231139](#)
18. Passamaneck YJ, Di Gregorio A. *Ciona intestinalis*: chordate development made simple. *Dev Dyn*. 2005; 233: 1–19. PMID: [15765512](#)
19. Davidson B, Christiaen L. Linking chordate gene networks to cellular behavior in ascidians. *Cell*. 2006; 124: 247–250. PMID: [16439196](#)
20. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*. 2006; 439: 965–968. PMID: [16495997](#)
21. Brown CD, Johnson DS, Sidow A. Functional Architecture and Evolution of Transcriptional Elements That Drive Gene Coexpression. *Science*. 2007; 317: 1557–1560. PMID: [17872446](#)
22. Stolfi A, Gainous TB, Young JJ, Mori A, Levine M, Christiaen L. Early chordate origins of the vertebrate second heart field. *Science*. 2010; 329: 565–568. doi: [10.1126/science.1190181](#) PMID: [20671188](#)
23. Abitua PB, Wagner E, Navarrete IA, Levine M. Identification of a rudimentary neural crest in a non-vertebrate chordate. *Nature*. 2012; 492: 104–107. doi: [10.1038/nature11589](#) PMID: [23135395](#)
24. Passamaneck YJ, Katikala L, Perrone L, Dunn MP, Oda-Ishii I, Di Gregorio A. Direct activation of a notochord *cis*-regulatory module by Brachyury and FoxA in the ascidian *Ciona intestinalis*. *Development*. 2009; 136: 3679–3689. doi: [10.1242/dev.038141](#) PMID: [19820186](#)
25. Kelkar YD, Strubczewski N, Hile SE, Chiaromonte F, Eckert KA, Makova KD. What Is a Microsatellite: A Computational and Experimental Definition Based upon Repeat Mutational Behavior at A/T and GT/AC Repeats. *Genome Biol Evol*. 2010; 2: 620–635. doi: [10.1093/gbe/evq046](#) PMID: [20668018](#)
26. Evans AL, Faial T, Gilchrist MJ, Down T, Vallier L, Pedersen RA, et al. Genomic Targets of Brachyury (T) in Differentiating Mouse Embryonic Stem Cells. *PLoS ONE*. 2012; 7: e33346. doi: [10.1371/journal.pone.0033346](#) PMID: [22479388](#)
27. Kubo A, Suzuki N, Yuan X, Nakai K, Satoh N, Imai KS, et al. Genomic *cis*-regulatory networks in the early *Ciona intestinalis* embryo. *Development*. 2010; 137: 1613–1623. doi: [10.1242/dev.046789](#) PMID: [20392745](#)
28. Kispert A, Koschorz B, Herrmann BG. The T protein encoded by *Brachyury* is a tissue-specific transcription factor. *EMBO J*. 1995; 14: 4763–4772. PMID: [7588606](#)
29. Casey ES, O'Reilly MA, Conlon FL, Smith JC. The T-box transcription factor Brachyury regulates expression of *eFGF* through binding to a non-palindromic response element. *Development*. 1998; 125: 3887–3894. PMID: [9729496](#)
30. Di Gregorio A, Levine M. Regulation of *Ci-tropomyosin-like*, a Brachyury target gene in the ascidian, *Ciona intestinalis*. *Development*. 1999; 126: 5599–5609. PMID: [10572037](#)
31. Conlon FL, Fairclough L, Price BM, Casey ES, Smith JC. Determinants of T box protein specificity. *Development*. 2001; 128: 3749–3758. PMID: [11585801](#)
32. Kusch T, Storck T, Walldorf U, Reuter R. Brachyury proteins regulate target genes through modular binding sites in a cooperative fashion. *Genes Dev*. 2002; 16: 518–529. PMID: [11850413](#)
33. Katikala L, Aihara H, Passamaneck YJ, Gazdoui S, José-Edwards DS, Kugler JE, et al. Functional Brachyury binding sites establish a temporal read-out of gene expression in the *Ciona* notochord. *PLoS Biol*. 2013; 11: e1001697. doi: [10.1371/journal.pbio.1001697](#) PMID: [24204212](#)

34. José-Edwards DS, Oda-Ishii I, Nibu Y, Di Gregorio A. Tbx2/3 is an essential mediator within the Brachyury gene network during *Ciona* notochord development. *Development*. 2013; 140: 2422–2433. doi: [10.1242/dev.094227](https://doi.org/10.1242/dev.094227) PMID: [23674602](https://pubmed.ncbi.nlm.nih.gov/23674602/)
35. Imai KS, Hino K, Yagi K, Satoh N, Satou Y. Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development*. 2004; 131: 4047–4058. PMID: [15269171](https://pubmed.ncbi.nlm.nih.gov/15269171/)
36. José-Edwards DS, Kerner P, Kugler JE, Deng W, Jiang D, Di Gregorio A. The Identification of Transcription Factors Expressed in the Notochord of *Ciona intestinalis* Adds New Potential Players to the Brachyury Gene Regulatory Network. *Dev Dyn*. 2011; 240: 1793–1805. doi: [10.1002/dvdy.22656](https://doi.org/10.1002/dvdy.22656) PMID: [21594950](https://pubmed.ncbi.nlm.nih.gov/21594950/)
37. Miwata K, Chiba T, Horii R, Yamada L, Kubo A, Miyamura D, et al. Systematic analysis of embryonic expression profiles of zinc finger genes in *Ciona intestinalis*. *Dev Biol*. 2006; 292: 546–554. PMID: [16519883](https://pubmed.ncbi.nlm.nih.gov/16519883/)
38. Brozovic M, Martin C, Dantec C, Dauga D, Mendez M, Simion P, et al. ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. *Nucleic Acids Res*. 2015.
39. Johnson DS, Zhou Q, Yagi K, Satoh N, Wong W, Sidow A. De novo discovery of a tissue-specific gene regulatory module in a chordate. *Genome Res*. 2005; 15: 1315–1324. PMID: [16169925](https://pubmed.ncbi.nlm.nih.gov/16169925/)
40. Kugler JE, Gazdoui S, Oda-Ishii I, Passamaneck YJ, Erives AJ, Di Gregorio A. Temporal regulation of the muscle gene cascade by Macho1 and Tbx6 transcription factors in *Ciona intestinalis*. *J Cell Sci*. 2010; 123: 2453–2463. doi: [10.1242/jcs.066910](https://doi.org/10.1242/jcs.066910) PMID: [20592183](https://pubmed.ncbi.nlm.nih.gov/20592183/)
41. Kusakabe T, Yoshida R, Ikeda Y, Tsuda M. Computational discovery of DNA motifs associated with cell type-specific gene expression in *Ciona*. *Dev Biol*. 2004; 276: 563–580. PMID: [15581886](https://pubmed.ncbi.nlm.nih.gov/15581886/)
42. Wang W, Christiaen L. Transcriptional enhancers in ascidian development. *Curr Top Dev Biol*. 2012; 98: 147–172. doi: [10.1016/B978-0-12-386499-4.00006-9](https://doi.org/10.1016/B978-0-12-386499-4.00006-9) PMID: [22305162](https://pubmed.ncbi.nlm.nih.gov/22305162/)
43. Haussler M, Jaszczyszyn Y, Christiaen L, Joly J-S. A *Cis*-Regulatory Signature for Chordate Anterior Neuroectodermal Genes. *PLoS Genet*. 2010; 6: e1000912. doi: [10.1371/journal.pgen.1000912](https://doi.org/10.1371/journal.pgen.1000912) PMID: [20419150](https://pubmed.ncbi.nlm.nih.gov/20419150/)
44. Markstein M, Zinzen R, Markstein P, Yee K-P, Erives A, Stathopoulos A, et al. A regulatory code for neurogenic gene expression in the *Drosophila* embryo. *Development*. 2004; 131: 2387–2394. PMID: [15128669](https://pubmed.ncbi.nlm.nih.gov/15128669/)
45. Jin H, Stojnic R, Adryan B, Ozdemir A, Stathopoulos A, Frasch M. Genome-wide screens for in vivo Tbx1 binding sites identify cardiac enhancers with diverse functional architectures. *PLoS Genet*. 2013; 9: e1003195. doi: [10.1371/journal.pgen.1003195](https://doi.org/10.1371/journal.pgen.1003195) PMID: [23326246](https://pubmed.ncbi.nlm.nih.gov/23326246/)
46. Rastegar S, Hess I, Dickmeis T, Nicod JC, Ertzer R, Hadzhiev Y, et al. The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. *Dev Biol*. 2008; 318: 366–377. doi: [10.1016/j.ydbio.2008.03.034](https://doi.org/10.1016/j.ydbio.2008.03.034) PMID: [18455719](https://pubmed.ncbi.nlm.nih.gov/18455719/)
47. Johnson DS, Davidson B, Brown CD, Smith WC, Sidow A. Noncoding regulatory sequences of *Ciona* exhibit strong correspondence between evolutionary constraint and functional importance. *Genome Res*. 2004; 14: 2448–2456. PMID: [15545496](https://pubmed.ncbi.nlm.nih.gov/15545496/)
48. Kim JH, Waterman MS, Li LM. Diploid genome reconstruction of *Ciona intestinalis* and comparative analysis with *Ciona savignyi*. *Genome Res*. 2007; 17: 1101–1110. PMID: [17567986](https://pubmed.ncbi.nlm.nih.gov/17567986/)
49. Doglio L, Goode DK, Pelleri MC, Pauls S, Frabetti F, Shimeld SM, et al. Parallel evolution of chordate *cis*-regulatory code for development. *PLoS Genet*. 2013; 9: e1003904. doi: [10.1371/journal.pgen.1003904](https://doi.org/10.1371/journal.pgen.1003904) PMID: [24282393](https://pubmed.ncbi.nlm.nih.gov/24282393/)
50. Jahangiri L, Nelson AC, Wardle FC. A *cis*-regulatory module upstream of *deltaC* regulated by Ntla and Tbx16 drives expression in the tailbud, presomitic mesoderm and somites. *Dev Biol*. 2012; 371: 110–120. doi: [10.1016/j.ydbio.2012.07.002](https://doi.org/10.1016/j.ydbio.2012.07.002) PMID: [22877946](https://pubmed.ncbi.nlm.nih.gov/22877946/)
51. Anno C, Satou A, Fujiwara S. Transcriptional regulation of *ZicL* in the *Ciona intestinalis* embryo. *Dev Genes Evol*. 2006; 216: 597–605. PMID: [16705435](https://pubmed.ncbi.nlm.nih.gov/16705435/)
52. Di Gregorio A, Corbo JC, Levine M. The regulation of *forkhead/HNF-3beta* expression in the *Ciona* embryo. *Dev Biol*. 2001; 229: 31–43. PMID: [11133152](https://pubmed.ncbi.nlm.nih.gov/11133152/)
53. Dunn MP, Di Gregorio A. The evolutionarily conserved *leprecan* gene: its regulation by Brachyury and its role in the developing *Ciona* notochord. *Dev Biol*. 2009; 328: 561–574. doi: [10.1016/j.ydbio.2009.02.007](https://doi.org/10.1016/j.ydbio.2009.02.007) PMID: [19217895](https://pubmed.ncbi.nlm.nih.gov/19217895/)
54. Tamplin OJ, Cox BJ, Rossant J. Integrated microarray and ChIP analysis identifies multiple Foxa2 dependent target genes in the notochord. *Dev Biol*. 2011; 360: 415–425. doi: [10.1016/j.ydbio.2011.10.002](https://doi.org/10.1016/j.ydbio.2011.10.002) PMID: [22008794](https://pubmed.ncbi.nlm.nih.gov/22008794/)

55. Jeong Y, Epstein DJ. Distinct regulators of *Shh* transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node. *Development*. 2003; 130: 3891–3902. PMID: [12835403](#)
56. Muller F, Chang B, Albert S, Fischer N, Tora L, Strahle U. Intronic enhancers control expression of zebrafish *sonic hedgehog* in floor plate and notochord. *Development*. 1999; 126: 2103–2116. PMID: [10207136](#)
57. Yagi K, Satou Y, Satoh N. A zinc finger transcription factor, ZicL, is a direct activator of *Brachyury* in the notochord specification of *Ciona intestinalis*. *Development*. 2004; 131: 1279–1288. PMID: [14993185](#)
58. Alten L, Schuster-Gossler K, Eichenlaub MP, Wittbrodt B, Wittbrodt J, Gossler A. A Novel Mammal-Specific Three Partite Enhancer Element Regulates Node and Notochord-Specific *Noto* Expression. *PLoS ONE*. 2012; 7: e47785. doi: [10.1371/journal.pone.0047785](#) PMID: [23110100](#)
59. Sawada A, Nishizaki Y, Sato H, Yada Y, Nakayama R, Yamamoto S, et al. Tead proteins activate the *Foxa2* enhancer in the node in cooperation with a second factor. *Development*. 2005; 132: 4719–4729. PMID: [16207754](#)
60. Kugler JE, Passamaneck YJ, Feldman TG, Beh J, Regnier TW, Di Gregorio A. Evolutionary conservation of vertebrate notochord genes in the ascidian *Ciona intestinalis*. *Genesis*. 2008; 46: 697–710. doi: [10.1002/dvg.20403](#) PMID: [18802963](#)
61. Smith RP, Riesenfeld SJ, Holloway AK, Li Q, Murphy KK, Feliciano NM, et al. A compact, *in vivo* screen of all 6-mers reveals drivers of tissue-specific expression and guides synthetic regulatory element design. *Genome Biol*. 2013; 14: R72. doi: [10.1186/gb-2013-14-7-r72](#) PMID: [23867016](#)
62. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Mol Cell*. 2002; 9: 279–289. PMID: [11864602](#)
63. Ikuta T, Yoshida N, Satoh N, Saiga H. *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc Natl Acad Sci USA*. 2004; 101: 15118–15123. PMID: [15469921](#)
64. Oda-Ishii I, Di Gregorio A. Lineage-Independent Mosaic Expression and Regulation of the *Ciona multi-*dom Gene in the Ancestral notochord. *Dev Dyn*. 2007; 236: 1806–1819. PMID: [17576134](#)